Hydrol. Earth Syst. Sci. Discuss., 5, S2643–S2646, 2009
www.hydrol-earth-syst-sci-discuss.net/5/S2643/2009/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Mapping model behaviour using Self-Organizing Maps" *by* M. Herbst et al.

M. Herbst et al.

Received and published: 7 March 2009

The authors greatly acknowledge the constructive work of the reviewers which allowed us to submit an improved and more comprehensible version of the original manuscript. In the following we briefly discuss their main issues of concern:

Rereree#3 suggests that the end of the introduction requires rewriting as it anticipates the main results of the study.

Authors: We agree that this is quite unusual. Consequently the final version of the paper will not contain the penultimate paragraph of Sect. 1.

Rereree#3 claims that "basically all" equations in the manuscript need more description and the symbols and units should be made explicit.

Authors: Great care was taken to provide a brief but yet complete description of the Signature Indices and the SOM algorithm. We admit that some rather short descriptions, especially in Sect. 2.3 could lend themselves to confusion (see also Referee#2). Minor, but hopefully effective, modifications and additions to Sect. 2.3, Sect. 2.4 as well as the end of Sect. 2.2 should clarify all problems with respect to the equations and the symbols. In this sense, e.g. the use of variable $\boldsymbol{x}$ and $\boldsymbol{y}$ as well as the composition of the training data set X is now stated more explicitly. In Tab. 1 as well as Sect. 2.1 the fact that all parameters are unit less factors is made more explicit.

Rereree#3 “How certain are the authors that the SOM has found global minima for the individual Signature Indices? Maybe SCE-UA can be run as well for the individual SI’s and the results can be compared to what the SOM has found. The SOM optimization is not completely free for the individual SI’s, and maybe that can influence the outcomes.“

Authors: The authors do not make a claim that the proposed SOM technique is capable of identifying the global minima of the individual Signature Indices (nor a Pareto set of solutions). There are two reasons for this that are also mentioned in the paper: In contrast to the SCE-UA algorithm, the SOM does not constitute an iterative algorithm which is capable of searching the parameter space with potentially infinite resolution. Instead the SOM is used to select a set of model realizations from a pre-defined set of model realizations obtained with NASIM as described in Sect. 2.1 and 2.2 (see also p. 3539, lines 15-18). Consequently, the SOM based “optimization” depends on a fixed sampling density in the parameter space. An optimization of single Signature Indices based on the SCE-UA algorithm would therefore always outperform the SOM approach. Further, the “BMU model realizations” are selected according to the criteria which are expressed in Eq. (9) and (14) of the paper: The reference vector (i.e. node) with the smallest Euclidean distance to a given vector of Signature Indices is selected. The model realizations that have been projected to this node in the course of the training constitute the solution provided by the SOM. It has been shown that the SOM is capable of providing a set of model realizations that satisfy a given set of cri-

HESSD

5, S2643–S2646, 2009

Interactive Comment

teria such that e.g. a good approximation to the FDC characteristics of the time series of observed discharges can be achieved. However, the authors clearly point out the shortcomings that could arise from the approach of minimizing the Euclidean distance in the five-dimensional Signature Index Space (p. 3539 line 24ff): “As to the aspect of using the SOM for multi-criteria optimization, it should be borne in mind that determining the BMU of the measured time series is equivalent to converting a multi-objective optimization to a single-criterion problem by means of weighting the objective function (Zadeh, 1963; Madsen, 2003). This might be of importance in cases where the multi-criteria Pareto front is non-convex; i.e. in general finding the BMU might not provide a genuine multi-criteria solution.”

As the SOM-BMU approach is used to solve a multi-criteria problem it is automatically “not completely free for the individual SI’s” because the solution to such a problem consists of a set of Pareto optimal model realizations that embody the trade-offs between the individual criteria. The solutions of single-criteria optimizations of the individual SI’s, instead, mark the extreme ends of a Pareto-front. Nevertheless, in the case of the SOM presented in our study, the authors found that Pareto-optimal SI data sets were projected to several nodes of a rather compact region on the map that also includes the BMU of the observed time series.

Referee#3: asks how the SOM methodology compares to other multivariate techniques as e.g. “PCA, RDA and CCA“ that also provide good tools for visualization of multi-dimensional data sets.

Authors: We gratefully acknowledge this valuable question which is addressed very shortly at the end of Sect. 2.3 (p. 3529, line 18ff) and shall be answered here in more detail following Kaski (1997).

Principal Component Analysis (PCA) can be used to determine the axes of a linear subspace that preserve a maximum of the variance of the data. However, as these axes are linear, PCA cannot account properly for curved or arbitrarily shaped data distributions which are frequently encountered e.g. when evaluating the results of Monte-Carlo experiments with environmental models. In conjunction with SOM PCA is often used to visualize the training process and to show how the reference vectors adapt to the distribution of the training data.

Principle curves (or CCA) constitute a generalized form of PCA. Instead of projecting the data into a linear subspace, the principle curves span a non-linear sub-space using smooth curves that are formed by the mean value of all data points that project to it. Thus, each point on a principal curve constitutes a conditional expectation value. SOM can be seen as essentially equivalent to a discretized form of principle curves as the reference vectors represent local conditional expectations of the data. This property, again, can be visualized in an impressing way when using PCA to display the position of the reference vectors and the input data items in a three-dimensional coordinate system.

HESSD

5, S2643–S2646, 2009

Interactive Comment